AstroShelf: Understanding the Universe through Scalable Navigation of a Galaxy of Annotations [1]

Panayiotis Neophytou, Roxana Gheorghiu, Rebecca Hachey, Timothy Luciani, Boyu Sun,
Alexandros Labrinidis, G. Elisabeta Marai, Panos K. Chrysanthis

Department of Computer Science, University of Pittsburgh

{panickos, roxana, reh59, tbl8, boyu, labrinid, marai, panos }@cs.pitt.edu

ABSTRACT

This demo presents AstroShelf, our on-going effort to enable astrophysicists to collaboratively investigate celestial objects from data originating from multiple sky surveys. The AstroShelf platform combines database and data stream, workflow and visualization technologies to provide means for querying and displaying telescope images (in a Google Sky manner), visualizations of spectrum data, and creating annotations. A key feature are Live Annotations which are the detection, ranking and delivery of events or annotations to users in real-time, based on their profile. We demonstrate how the user interaction, as well as the inner workings of our monitoring platform and ranking algorithms.

1. INTRODUCTION

The growing onslaught of astronomical data available presents a great challenge. Astronomy lacks an easy-to-use and scalable way to collect and distribute expert information about objects from datasets of tens of thousands to billions of individual events and objects. Over the next decade, the amount of information available to the typical astronomer will grow by two orders of magnitude both in raw data size and in the number of objects.

We have been working with a group of astrophysicists to develop *AstroShelf* which will enable them to collaboratively annotate sky objects and phenomena, as well as visualize parts of the sky using different algorithms.

AstroShelf pursues two research directions, each of which has the potential to transform how astronomers interface with large datasets: (1) a *scalable annotation framework* to enable linking of observations to specific experiments, models, or other observations; and (2) a *continuous workflow enactment system* that would support automated reasoning in the presence of uncertainty.

In this demo we will demonstrate the design and usage of our Astroshelf prototype system, and show how it is being used in two timely scientific programs: (a) methods to incorporate images and catalogs from disparate datasets, using multi-wavelength data from two surveys, allowing astronomers to study how the demographics of galaxies have changed over the last 8 billion years; and (b) techniques for rapid classification of transient phenomena, dissemination of their properties, and incorporation of feedback from follow-up observations, using properties of time-variable events found in one survey.

This project enables astrophysicists to share information and expert opinions quickly and widely, about each new observation or event, fundamentally changing our ability to learn about the Universe; such functionality can also be applied in support of other scientific domains. A similar effort has been demonstrated in [1], where the system was displaying earthquake information. That system, though, did not address collaboration through annotations.

We will demonstrate the following features of AstroShelf:

- The AstroShelf's User Interface that supports: (1) search and visualization of sky objects, their data and meta-data, (2) annotations management, and (3) specification of interest and live annotations.

- The use of CONFLuEnCE [6] to support the monitoring of data streams which are either entered the system as new celestial events (e.g., new supernova detected) or are created within the system as new annotations. The set of notifications from detected events are ranked according to each user's profile.

The current version of AstroShelf is being packaged to be used as an educational tool, allowing astronomy students to collaboratively explore the sky.

2. ASTROSHELF

AstroShelf's main modules are a user interface (dashboard) with the ability to display sky images and generally enable user interaction with the system, an annotations management system and a monitoring module for real-time processing of annotations and sky update events.
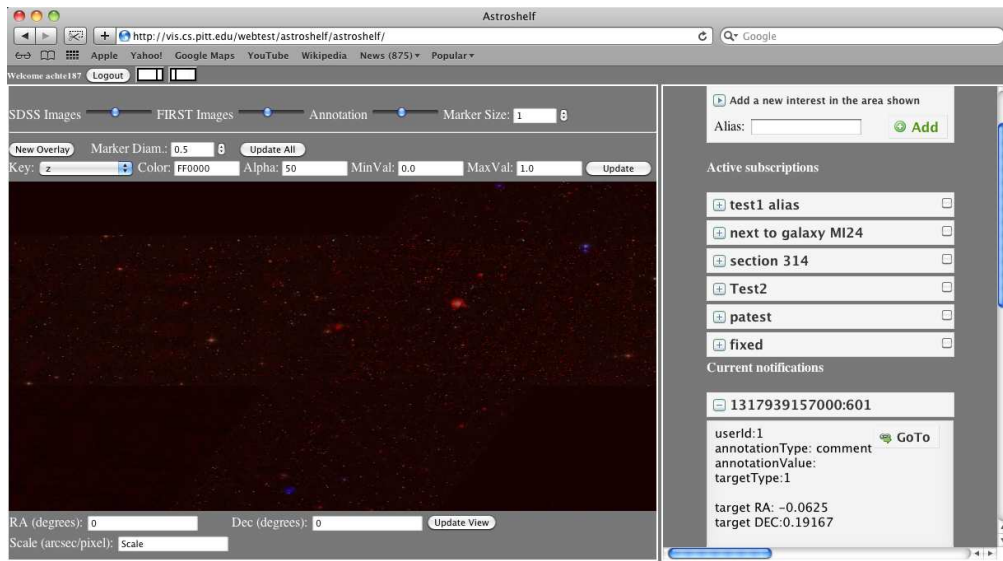
2.1 AstroShelf User Interface

AstroShelf's User Interface (UI), depicted in Figure 1 provides the user with a *SkyView* component, which displays sky images from various surveys on multiple layers. Meta-data regarding celestial objects are also displayed, as annotations, in another layer. The UI also facilitates searching of objects and annotations on objects or areas, as well as the ability to add new annotations. As part of the visualization of meta-data the interface allows the user to generate *trend images* on objects retrieved using the search mechanism. Lastly, a *Live annotations* module provides real-time updates on the meta-data, new annotations, or new transient events. We elaborate these user interface components next.

2.1.1 SkyView

The SkyView in our system is an interactive, 2D HTML5 canvas, implemented in webGL. Within this canvas the user can interact with data and analyze the output of other components of our platform. The images of the sky are provided by astronomical surveys such as SDSS (www.sdss.org) and FIRST(sundog.stsci.edu),

Figure 1: Screenshot from AstroShelf's User Interface, currently showing the SkyView (left) and the Live Annotations View (right).

as well as simulations such as LSST(www.lsst.org) and are displayed using different levels of opacity. The SkyView also allows zooming and panning. All images are stored in a spatial index tree, implemented as an S2-tree[8]. Finally, SkyView uses pixel-based annotations to view data provided by our Annotations Management System as well as survey databases. Pixel-based annotations allow the user to visualize thousand of objects in a single image, possibly spotting trends in the data otherwise impossible to perceive (e.g., markers for heavily annotated objects appear brighter than those of lightly annotated objects). Since these annotations are pixel-sized, numerous queries can be viewed at once without hindering the overall view.

2.1.2 Trend Images

Trend images are used to visualize overall patterns in a data set. A trend image consists of a set of pixels, each mapped to a specific data point in such a way that patterns in the data become apparent from patterns in the resulting image.

An example of how a trend image could be used is to visualize patterns in spectroscopic data from astronomy observations. Spectroscopic data involves measuring the flux or intensity of light at different wavelengths for objects of interest. The trend image in Figure 2 consists of spectroscopic data for 139 Type Ia supernovae from Balland et al. [2]. Each horizontal row corresponds to the spectra for a supernova, and they are ordered from top to bottom in increasing redshift. Along the x-axis the data points are ordered in increasing wavelength from left to right. The grayscale pixel values correspond to the flux of the object at that wavelength. The general broad features of the Type Ia supernovae can be seen in the dark and light bands that represent characteristic features of explosions. These smoothly trace back over redshift indicating that these are a consistent class of an object.

AstroShelf supports generating a trend image for any number and type of objects specified by the user. In the current demo, the objects are from the specObj view of the seventh data release of SDSS but, in general, AstroShelf can support any catalogue of celestial data. Closer inspection of the spectra of outliers is supported by the clickable image map which links a graph of the spectral data of a single object to the horizontal line of pixels it corresponds to.
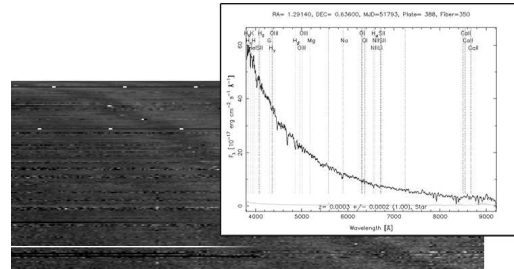
Figure 2: Trend Image for 139 Type Ia supernovae.

2.1.3 Live Annotations View

The Live Annotations module of the user interface displays, in real-time, any new notifications regarding areas and types of objects which are of interest to the current user. The user may register multiple *interests* with the system and the updates will reach her by the preferred means (e.g., browser UI when logged in, email when not, notifications on mobile phone). The user may filter the notifications by interest, and then by clicking on them she will be navigated to the area of the sky where the annotation was made (through the SkyView).

The Live Annotations view not only displays the new annotations made in specific areas and specific objects, but it also provides the user with trending areas or objects which are relevant to the user's interests. This proves to be particularly useful when the number of notifications is either high, or the user is registered to areas of the sky where there is not too much activity, while in other areas (unbeknownst to the user), there has been a new discovery that may be of interest to the user.

2.2 Scalable Annotations Framework

The Annotations Service is the system component which handles annotations on sky objects or sky locations in varying granularity levels. The service also works as a notification service; it sends notifications of changes in data and meta-data to registered entities. The service exposes a RESTful API which can be used by third parties to insert, query or update annotations, given the right credentials.

In our system, any annotation is represented as a pair *<annotation type, annotation value>*. Because of that, our system supports any type of annotation one may want to add (e.g., a tag, any text, link to a file with experimental results, image, link to a paper, provenance information, etc.) Annotations can be made on three different levels of granularity:

- Annotations on a *celestial object*. When the object already exists in our database, annotations are directly linked to that object. If the object does not exist then the user is able to create her personal survey and the new object definition becomes part of that survey. When visualizing the annotation on the SkyView, it appears on the object's RA/DEC coordinates.

- Annotations on a *set of objects* can be made by specifying a set of objects having some specific properties (based on a user defined query or view). Moreover, the system allows the definition of annotations on specific RA/DEC ranges. This can be seen as an annotation that is added on an area of the sky instead of particular objects, and any objects which may be added later in that area are also annotated with the area's annotation in a similar fashion as in view-based annotations.

- *Annotations on annotations*, can be made, for example, to represent revisions on annotation values or discussions between users on some annotation value.

Users can express their preferences using the AstroShelf interface. Preferences are considered to be in the form of positive or negative opinions towards an annotation. Using this valuable information, the system will assign a weight to each annotation that matches a user's query. This weight will express the intensity of interest in that particular annotation. Every annotation that was marked as *"preferred"* by a user will have a weight equal to 1 for that particular user. All other annotations will have a weight value between 0 and 1. The system then suggests possible other interesting objects and annotations if the weight is above some threshold.

When assigning the weights, the system tries to find relevant similarities between two objects. These similarities can refer to meta-data information (i.e., position on the sky, type of object, type of event) or to annotations attached to that particular objects. For example, if two objects, A and B, are linked through annotations to the same paper, when users' query output returns A, the system may suggest as relevant object B with all annotations attached to it.

2.3 CONFLuEnCE

As mentioned in the introduction, all the interactions in AstroShelf are facilitated by a workflow engine. Specifically, CONFLuEnCE, short for CONtinuous workFLow ExeCution Engine, is the implementation of our Continuous Workflow (CWf) model [7, 5]. It was built on top of Kepler [4], and can facilitate both scientific and business reactive, monitoring, and collaborative applications.

In the context of supernova and galaxy classification, we have designed a continuous workflow to run on CONFLuEnCE and we plan to demonstrate its efficiency during the presentation. The goal of this workflow is to monitor the activity of inserting, updating or deleting annotations as well as integrating the detection of transient events from various sky surveys that are of interest to the users, all in real-time. After processing these events the workflow will notify the users about them, and it will receive their feedback. By interacting with the workflow, the users may refine the annotations, iterating over them until they reach a consensus.

3. INTERACTIONS

The system interactions and flow of events are as follows (numbered as in Figure 3):

1) Using the AstroShelf UI users can define and name areas in the sky that are of interest to them. Additionally they may define the type of events of interest(e.g., new annotation, new supernova, galaxy classification, etc.) This expression of interest is pushed to CONFLuEnCE and is registered into an R-Tree spacial index which resides inside the actor "Tag Interest".

2) Using the SkyView, users can annotate objects, groups of objects or arbitrary points in the sky with any information they deem important to share. Classifications of galaxies or supernovae made by the users through the UI are also recorded as annotations All of these annotations are inserted into the Annotations Engine through a specialized API.

3) Any new annotation is detected by the Event Reporting module that directly forwards it to the continuous workflow on CONFLuEnCE. Other types of events pushed to the monitoring workflow are transient events detected by various sky surveys, which are also available through an aggregation service, called SkyAlert (skyalert.org).

4) Once the aforementioned events enter the system they are tagged by the "Tag Interest" actor with the user ids of those who previously expressed interest in the area and type of the object attached to the annotation (Interaction 1, Fig 1). Then they are filtered depending on the event type, and follow different paths in the workflow.

   a. In supernovae events, the supernova object is matched with its host galaxy and this matching is verified by the user through the browser interface and then calculate the redshift probability distribution.

   b. All other events are joined with data available from various external catalogs. This information will help the users when they provide feedback about an annotation.

5) Once all the necessary data have been attached to the data objects, users tagged on those objects are notified directly on their browser or through email, SMS, twitter etc.

6) The notified users use the AstroShelf interface to express their opinion on the annotations. The opinions are tagged as positive or negative and split accordingly. The "Split neg/pos" step groups the opinions according to the object id and the sentiment of the opinion. The window size of the group-by is time-based to measure the temporal density of each opinion.

7) At this step the workflow evaluates the overall consensus on the various opinions (positive or negative). It will then create another annotation on the object that captures the consensus. This new annotation goes back into the workflow and the cycle continues.

8) In this step the system uses the Annotations Service to decide whether or not this event might interest the user. It does this using a remote procedure call to the recommendation system, which uses user's profiles (described in Section 2.2).

9) The "Time-based windowed ranking" actor, groups events by the user ids. It uses the time-based windows (e.g., every 5 minutes, size 30 minutes) and ranks the recommendations for each user, again using profiling information from the Annotation Service.

As it can be seen from the steps described above, the process of annotating sky objects is a loop which runs until the users collaboratively converge to a significant opinion.
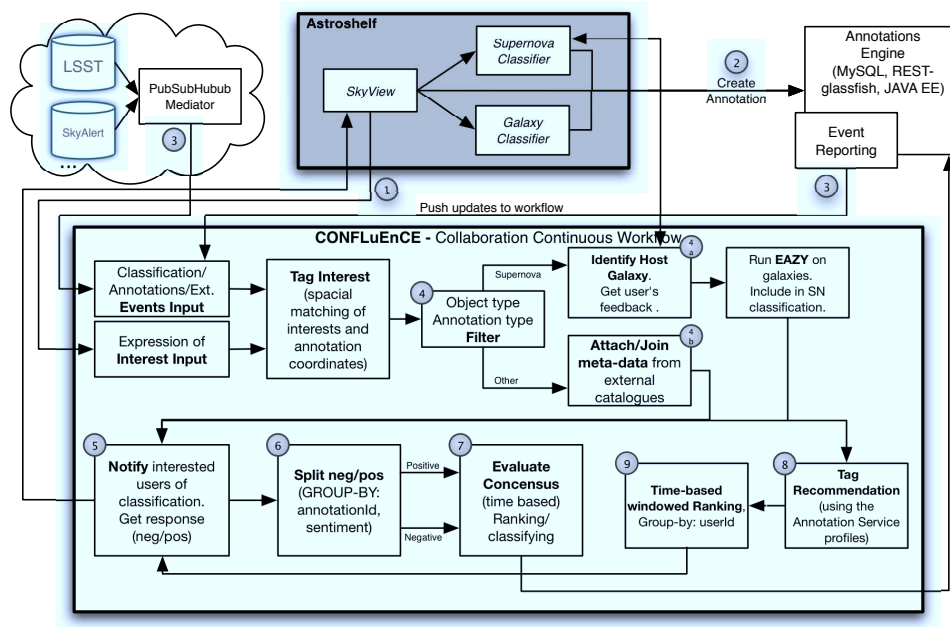
Figure 3: High-level design of the AstroShelf collaboration platform.

4. DEMONSTRATION

In our demonstration we would like our audience to interact with the AstroShelf system as users of our platform. They will be able to use multiple displays to enter new annotations, express interest in areas in the sky and then get notified by various ways (e.g., on screen, email etc.) when new events of their interest are detected. In order to make it interactive we assume all annotations are public and every user can see any annotation. In this way, a user can define an area of interest and any annotation added to that area will be pushed, as a notification, to their account.

Another aspect in our project is the visualization tool. One can see the image of any portion of the sky, zoom in for a closer look of celestial objects and surrounding area and make annotations. In the visualization environment users can play with different ways of sky visualization. They can see only the redshift, (e.g. the light coming from an object that increases proportionally with the wavelength), only the image of the sky as found in different astronomy surveys or both of them combined. Moreover, the interface allows users to see a mark on the sky next to objects that have been annotated already. Each mark has a different color intensity depending on how many annotations are defined for that particular patch of sky.

Finally, we will demonstrate to our audience how the AstroShelf system can be used through an API. By eliminating the need of a user interface, we allow our users to write and retrieve annotations in a fast way. We will show how easy it is to submit multiple annotations in the same time and how annotations can be fetched from our system. This will be demonstrated using automated scripts, which will also help in demonstrating the notifications and recommendations system, since the scripts will generate enough traffic to trigger the recommendations mechanism.

In the end, we would like to demonstrate that our system addresses the following problems:

- *Annotations*: We have two different ways of interactively adding annotations (i.e., using the visualization tool or programmatically, using the annotation service). In this way, we address two different problems: a) an interface that may become a bottleneck in the process of adding annotations by hand and b) a system that can become too technical for someone that only wants to visualize the sky and add various notes on things she discovers.

- *Workflows*: Our system provides an interactive way of staying connected with the community through a complex continuous workflow execution engine. Annotations relevant to users preferences are pushed continuously and can be accessed on different platforms.

- *Visualization*: AstroShelf visualization system is a great way to look at the sky images and celestial objects. Using the various layers AstroShelf system offers (i.e., SkyView, Redshift slider, Trend images) astrophysicists can interact directly with the sky images and discover faster interesting areas or objects on the sky.

5. REFERENCES

[1] M. Ali et al. Online visualization of geospatial stream data using the worldwide telescope. In *p VLDB*, 4:1379–1382, 2011.

[2] C. Balland et al. The ESO/VLT 3rd year Type Ia supernova data set from the supernova legacy survey. In *A&A*, 507:85–103, 2009.

[3] G. B. Brammer et al. Eazy: A fast, public photometric redshift code. *The Astrophysical Journal*, 686(2):1503, 2008.

[4] B. Ludäscher et al. Scientific workflow management and the kepler system. *Concurrency and Computation: Practice and Experience*, 18(10):1039–1065, 2006.

[5] P. Neophytou, P. Chrysanthis, A. Labrinidis. Towards continuous workflow enactment systems. In *CollaborateCom*, pp. 162–178, 2008.

[6] P. Neophytou, P. Chrysanthis, A. Labrinidis. Confluence: Continuous workflow execution engine. In *SIGMOD*, pp. 1311–1314, 2011.

[7] P. Neophytou, P. Chrysanthis, A. Labrinidis. Confluence: Implementation and application design. In *CollaborateCom*, 2011.

[8] H. Wan, C.-S. Perng. The S2-Tree: An index structure for subsequence matching of spatial objects. In *PAKDD*, pp. 312–323, 2001.